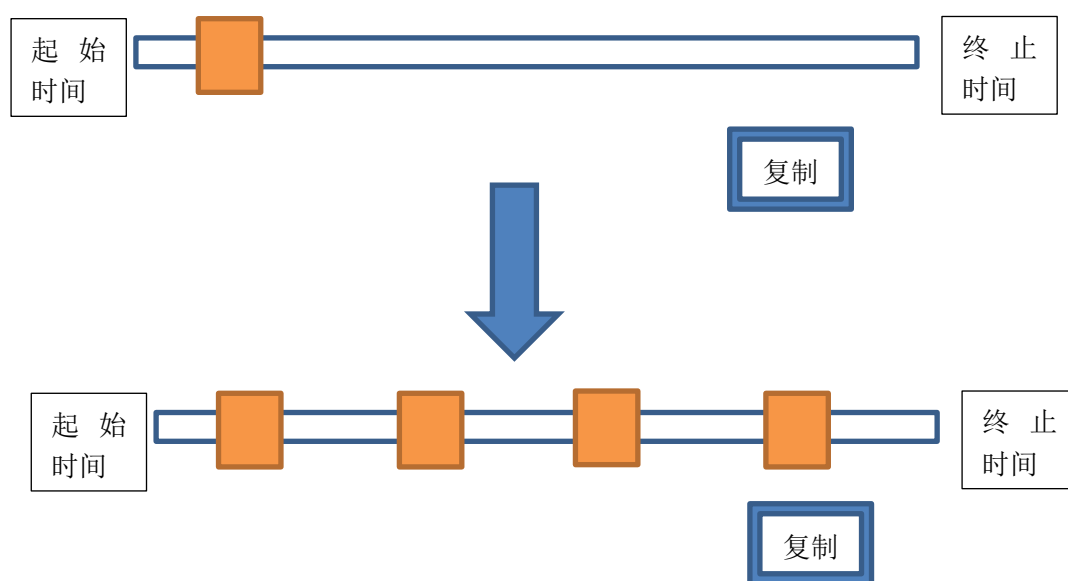


TVCG 论文思考

1、时间复制（微创新）

时空数据中时间上的选择是交互中的重要问题。在时空数据分析中，分析师常需要进行多时间段的数据比较，例如每周五的车辆轨迹、每天高峰时段的车流等。现有的时间交互大都是只是单尺度的时间选择，比如按天、日或小时等。对于多尺度时间选择，一方面难以分段，另外交互上需要多次选择较为麻烦。

我们的设计试图通过时间复制实现多尺度的时间选择，也就是说，用户选择一个时间段后直接复制，就可以在整个时间线上找出对应的多个相同时间段的时间片段，如下图所示。图中起始时间和终止时间可以设定，用户选择一个时间段（黄色框），点击复制，则自动在时间线上生成同周期时间。



2、哈希函数

在前面的论文中并没有讲述哈希函数的选择，对于长字节的数据，哈希函数生成的 key 值要尽量短，而且冲突率小。最近找到一个比较好的哈希函数 MurmurHash，从作者公布的测试结果看可以满足我们的要求。

3、预读机制（创新）

对于数据探索而言，无法将所有数据加载到内存。常见的数据探索方法有两种：采样部分数据展示近似结果；预读当前数据时空相邻数据。对于前者而言，当进行大规模聚集计算时，这种方法可以给用户较好的提示作用让用户了解总体的数据模式。但是，对于道路查询而言，这种查询是局部的，本身的数据量就不大，采样后准确性难以保证。根据当前查看数据的特点事先将附近的数据读取到内存，这样可以减少交互上的延时。现有的预读机制有根据视点位置预读、根据时间点预读等。然而面向数据可视化的数据预读机制研究较少。在我们的应用中，可以根据不同的查询要求由用户选择预读机制。可以提供根据当前查询位置、当前查询时间以及周期查询时间等预读机制。

VLDB 论文思考

1、题目：面向轨迹的城市数据可视化查询

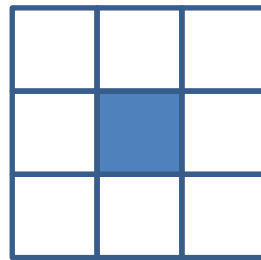
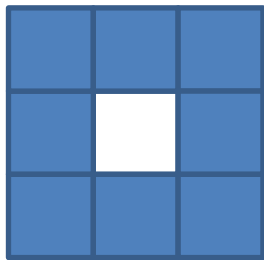
2、查询模型

模型中设定三方面的约束：空间、时间、关系。关系可以由用户根据数据特征定义。在论文中，我们给定两种关系定义方法：

1) 出租车轨迹：定义车流与道路之间的拓扑关系

2) 人群移动数据：定义人流的聚集/消散关系

对于第一种关系已经实现，不再描述。对于第二种关系，其实可以通过 LISA 统计值计算。该值定义的是局部区域内相同统计值的聚集情况，如果在指定的区域上，其周边区域的值都比该区域上的值高则用 1 表示，反之用 -1 表示，如果相等则用 0 表示。



3、查询算法

算法上主要考虑数据量较大时，在单个的空间单元格上的数据量极大，难以快速找到指定时间范围内的数据。按时间段将数据划分到不同机器上后，可以分而治之地查询各个时间段上的数据，较少查询时间。

4、应用

出租车轨迹的应用主要是进行交通状况的检测，而对于人群数据的应用目前想到的是聚散情况。对于聚散分析，已有工作的分析只能在局部区域判断聚散模式，但是无法知道聚散人群的来源和去向。根据查询的轨迹，可以判断人群轨迹的起点和终点，从而推断聚散的原因。